Our past interrogation of the Voynich Manuscript has deconstructed its esoteric symbols into a form more suitable for our ends, subjected its statistical properties to comparison with more mundane texts, and unearthed its hidden internal structures via the esoteric process of topic modelling. In this final post, we will build on the structures revealed in earlier posts to ask how, if at all, the Voynich Manuscript’s textual properties *shift* within the text itself. Are there significant discontinuities in the writing, indicating a separation of the manuscript into meaningful sections? Or is the text merely a homogenous mass more suggestive of a rote, mechanical, generative procedure?

To address this question we will delve once more into the arcana of machine learning, and draw out the technique of *changepoint analysis*. This procedure aims to identify one or more points in a series of observations at which the underlying process that generates the data has somehow altered.

Once more, we will operate within the warm embrace of Bayesian statistics and exploit the Stan modelling language as our means to cast light into the darkness.

Reproducing, with only slight adaptations, the original Poisson example, our full joint probability would be:

$$p(\theta\_e, \theta\_l, c, \Omega) =  
p(\theta\_e)p(\theta\_l)p(c)p(\Omega|\theta\_e,\theta\_l,c)$$

Marginalising out \(c\), this can be represented as:

$$\begin{eqnarray}  
p(\Omega|\theta\_e, \theta\_l) &=& \sum\_{c=1}^Tp(c,\Omega|c,\theta\_e,\theta\_l)\\  
&=& \sum\_{c=1}^Tp(c)p(\Omega|c,\theta\_e,\theta\_l)  
\end{eqnarray}$$

The result is that our Stan model can be constructed by sampling across values of \(\theta\_e\) and \(\theta\_l\) for all possible values of \(c\). Due to the requirement to sum across all possible values of the discrete paramete, however, this subterfuge of marginalisation is restricted in general to *bounded* discrete parameters.

We can distill the above into a Stan model as given below.

Multinomial changepoint model

multinomial\_changepoint.stan

data {

int num\_obs; // Number of observations (rows/pages) in data.

int num\_cats; // Number of categories in data.

int y[num\_obs, num\_cats]; // Matrix of observations.

vector[num\_cats] alpha; // Dirichlet prior values.

}

transformed data {

// Uniform prior across all time points for changepoint.

real log\_unif;

log\_unif = -log(num\_obs);

}

parameters {

// Two sets of parameters.

// One (early) before changepoint, one (late) for after.

simplex[num\_cats] theta\_e;

simplex[num\_cats] theta\_l;

}

transformed parameters {

// // This code shows a slower, but easier to understand updating of log posterior via summation.

// vector[num\_obs] lp;

// lp = rep\_vector(log\_unif, num\_obs);

// for (s in 1:num\_obs)

// for (t in 1:num\_obs)

// lp[s] = lp[s] + multinomial\_lpmf(y[t,] | t < s ? theta\_e : theta\_l);

// This approach relies on dynamic programming to reduce runtime from quadratic to linear in num\_obs.

// See

vector[num\_obs] log\_p;

{

vector[num\_obs + 1] log\_p\_e;

vector[num\_obs + 1] log\_p\_l;

log\_p\_e[1] = 0;

log\_p\_l[1] = 0;

for( i in 1:num\_obs ) {

log\_p\_e[i + 1] = log\_p\_e[i] + multinomial\_lpmf(y[i,] | theta\_e );

log\_p\_l[i + 1] = log\_p\_l[i] + multinomial\_lpmf(y[i,] | theta\_l );

}

log\_p =

rep\_vector( -log(num\_obs) + log\_p\_l[num\_obs + 1], num\_obs) +

head(log\_p\_e, num\_obs) - head(log\_p\_l, num\_obs);

}

}

model {

// Priors

theta\_e ~ dirichlet( alpha );

theta\_l ~ dirichlet( alpha );

target += log\_sum\_exp( log\_p );

}

generated quantities {

simplex[num\_obs] changepoint\_simplex; // Simplex of locations for changepoint.

// Convert the log posterior to a simplex.

changepoint\_simplex = softmax( log\_p );

}

To launch the model in gentler waters, we apply it to Martorel and de Galba’s *Tirant lo Blanc* to discover where, if at all, de Galba’s major contributions to the text begin, and see the extent to which our Stan rendering reproduces the results of Riba and Ginebra’s analysis. As in their work we will focus only on those words with length shorter than 5 letters, where the key *stylistic* difference between authors makes itself apparent.

We place a relative uninformative [Dirichlet prior](https://en.wikipedia.org/wiki/Dirichlet_distribution) on the multinomial \(\theta\_e\) and \(\theta\_l\) parameters, with the vector of \(\alpha\) values set to one. This results in a uniform distribution across all possible simplexes. Note that this does *not* push the multinomial towards a uniform simplex, but instead that all possible simplexes are equally likely. \(\alpha \ge 1\) produces increasingly uniform simplexes; \(0 \le \alpha \le 1\) produces simplexes in which the probability mass is more likely to be concentrated in some given element.

With the resulting data and model in unison, we can see the results of this analytic process.

Tirant lo Blanc Changepoint |

Our multinomial automaton suggests a location for a significant stylistic changepoint in *Tirant lo Blanc* with the main concentration of probability mass around chapter 374. Perhaps unsurprisingly, but pleasingly, this is in close agreement with the earlier analysis of Riba and Ginebra, who placed their estimates between chapters 371 and 382.

*Tirant lo Blanc* analysis code

library( tidyverse )

library( tidyselect )

library( magrittr )

library( rstan )

library( tidytext )

# Load Tirant data

message( "Reading raw Tirant data..." )

tirant\_tbl <-

read\_csv( "data/tirant\_raw.csv", col\_names=FALSE ) %>%

rename( chapter = X1, text = X2 ) %>%

mutate( page = as.numeric( rownames(.) ) )

# Tokenize

tirant\_words <-

tirant\_tbl %>%

unnest\_tokens( word, text )

# Pivot the data wider to be presented to Stan as a matrix of multinomial samples.

tirant\_lengths <-

tirant\_words %>%

mutate( word\_length = str\_length( word ) ) %>%

mutate( word\_length = ifelse( word\_length > 9, 10, word\_length )) %>%

group\_by( page, word\_length ) %>%

summarise( count = n( )) %>%

pivot\_wider( names\_from = word\_length, values\_from = count ) %>%

ungroup %>%

select( -c(page,"5","6","7","8","9","10") ) %>%

select(sort(peek\_vars())) %>%

replace( is.na(.), 0 )

if( not( file.exists( "work/multinomial\_changepoint\_tirant\_fit.rds" ) ) ) {

message( "Fitting multinomial model.")

tirant\_multinom\_fit <-

stan( "multinomial\_changepoint.stan",

data=list(

num\_obs=487,

num\_cats=4,

y = as.matrix( tirant\_lengths ),

alpha = rep( 1, 4 ) ),

iter=16000,

control=list(

adapt\_delta=0.98,

max\_treedepth=15 ) )

saveRDS( tirant\_multinom\_fit, "work/multinomial\_changepoint\_tirant\_fit.rds" )

} else {

message( "Loading saved multinomial model.")

tirant\_multinom\_fit <- readRDS( "work/multinomial\_changepoint\_tirant\_fit.rds" )

}

# Plot the calculated changepoint probabilities.

# ('changepoint\_simplex').

mean\_changepoint\_prob <-

extract( tirant\_multinom\_fit )$changepoint\_simplex %>%

as\_tibble( .name\_repair="unique" ) %>%

summarise\_all( mean ) %>%

pivot\_longer( everything() ) %>%

rowid\_to\_column()

# Save values for plotting

saveRDS( mean\_changepoint\_prob, file="work/mean\_changepoint\_prob\_tirant.rds" )

*Tirant lo Blanc* data file

[*Tirant lo Blanc* formatted raw text data.](https://www.weirddatascience.net/wp-content/uploads/2020/02/tirant_raw.csv)

*Tirant lo Blanc* changepoint plot code

multinomial\_changepoint\_tirant\_plot.r

library( tidyverse )

library( magrittr )

library( ggthemes )

library( showtext )

library( grimoire ) #

# Fonts

font\_add( "main\_font", "resources/fonts/alchemy/1651 Alchemy/1651AlchemyNormal.otf")

font\_add( "bold\_font", "resources/fonts/alchemy/1651 Alchemy/1651AlchemyNormal.otf")

showtext\_auto()

mean\_changepoint\_prob <-

readRDS( "work/mean\_changepoint\_prob\_tirant.rds" )

changepoint\_plot <-

ggplot( mean\_changepoint\_prob ) +

geom\_col( aes( x=rowid, y=value ), fill=weird\_colours["blood"] ) +

labs( x="Chapter", y="Probability of Changepoint" ) +

theme(

panel.background = element\_rect(fill = "transparent", colour = "transparent"),

plot.background = element\_rect(fill = "transparent", colour = "transparent"),

plot.title = element\_text( family="bold\_font", colour=weird\_colours["ink"], size=22 ),

plot.subtitle = element\_text( family="bold\_font", colour=weird\_colours["ink"], size=12 ),

axis.text = element\_text( family="bold\_font", colour=weird\_colours["ink"], size=12 ),

axis.title.x = element\_text( family="bold\_font", colour=weird\_colours["ink"], size=12 ),

axis.title.y = element\_text( family="bold\_font", angle=90, colour=weird\_colours["ink"], size=12 ),

axis.line = element\_line( colour=weird\_colours["ink"] ),

panel.grid.major.x = element\_blank(),

panel.grid.major.y = element\_blank(),

panel.grid.minor.x = element\_blank(),

panel.grid.minor.y = element\_blank()

)

# grimoire::decorate\_plot() from

parchment\_plot <-

decorate\_plot(

title="Tirant lo Blanc Chapter Changepoint Probability",

subtitle="http://www.weirddatascience.net | @WeirdDataSci",

plot=changepoint\_plot,

bg\_image="resources/img/parchment.jpg",

footer="Data: http://einesdellengua.com/tirantloweb/tirantloblanch.html" )

save\_plot("output/multinomial\_changepoint\_tirant\_plot.pdf",

parchment\_plot,

base\_width = 16,

base\_height = 9,

base\_aspect\_ratio = 1.78 )

**Nos et mutamur in illis**

The changepoint model realises the dark imaginings of our predecessors. What horrors might it reveal when applied to the Voynich Manuscript? Following the logic of the above analysis, we will focus initially on the shorter words in the Voynich corpus.

To improve model fit we will place a little more information in the prior, setting the \(\alpha\) values for the Dirichlet prior to 0.6 to push the multinomial towards more concentrated probabilities. We also combine all single- and two-letter Voynich terms into a single category due to the small number of words falling into these categories[3](https://www.weirddatascience.net/2020/02/21/illuminating-the-illuminated-part-four-tempora-mutantur-changepoint-analysis-of-the-voynich-manuscript/#easy-footnote-bottom-3-2009).

The model can now reveal where, if at all, a likely fracture resides in the textual assemblage of the Voynich Manuscript.

Voynich Manuscript Changepoint |

Voynich Manuscript word length analysis code

multinomial\_changepoint\_voynich.r

library( tidyverse )

library( tidyselect )

library( magrittr )

library( rstan )

library( tidytext )

# Load Voynich data

message( "Reading raw Voynich data..." )

voynich\_tbl <-

read\_csv( "data/voynich\_raw.csv", col\_names=FALSE ) %>%

rename( folio = X1, text = X2 )

# Tokenize

voynich\_words <-

voynich\_tbl %>%

unnest\_tokens( word, text )

# Calculate the lengths of words

voynich\_pure\_lengths <-

voynich\_words %>%

transmute( word\_length = str\_length( word ) )

voynich\_pure\_lengths$count <- 1

# Pivot the data wider to be presented to Stan as a matrix of multinomial samples.

voynich\_lengths <-

voynich\_words %>%

mutate( word\_length = str\_length( word ) ) %>%

mutate( word\_length = ifelse( word\_length > 8, 9, word\_length )) %>%

mutate( word\_length = ifelse( word\_length < 2, 2, word\_length )) %>%

group\_by( folio, word\_length ) %>%

summarise( count = n( )) %>%

pivot\_wider( names\_from = word\_length, values\_from = count ) %>%

ungroup %>%

select( -c("folio", "5", "6", "7", "8", "9" )) %>%

select(sort(peek\_vars())) %>%

replace( is.na(.), 0 )

if( not( file.exists( "work/multinomial\_changepoint\_voynich\_fit.rds" ) ) ) {

message( "Fitting multinomial model.")

voynich\_seed <- 1912

num\_cats <- ncol( voynich\_lengths )

voynich\_multinomial\_fit <-

stan( "multinomial\_changepoint.stan",

data=list(

num\_obs=226,

num\_cats=num\_cats,

y = as.matrix( voynich\_lengths ),

alpha = rep( 0.6, num\_cats ) ),

chains=4,

iter=8000, seed=voynich\_seed,

control = list( adapt\_delta=0.99,

max\_treedepth=12 ) )

saveRDS( voynich\_multinomial\_fit, "work/multinomial\_changepoint\_voynich\_fit.rds" )

} else {

message( "Loading saved multinomial model.")

voynich\_multinomial\_fit <- readRDS( "work/multinomial\_changepoint\_voynich\_fit.rds" )

}

# Plot the calculated changepoint probabilities.

# ('changepoint\_simplex').

mean\_changepoint\_prob <-

extract( voynich\_multinomial\_fit )$changepoint\_simplex %>%

as\_tibble( .name\_repair="unique" ) %>%

summarise\_all( mean ) %>%

pivot\_longer( everything() ) %>%

rowid\_to\_column()

# Plot the calculated changepoint probabilities.

# ('changepoint\_simplex').

mean\_log\_p <-

extract( voynich\_multinomial\_fit )$log\_p %>%

as\_tibble( .name\_repair="unique" ) %>%

summarise\_all( mean ) %>%

pivot\_longer( everything() ) %>%

rowid\_to\_column()

# Save mean values for plotting

saveRDS( mean\_changepoint\_prob, file="work/mean\_changepoint\_prob\_voynich.rds" )

Voynich Manuscript data file

[Voynich Manuscript formatted raw text data.](https://www.weirddatascience.net/wp-content/uploads/2020/02/voynich_raw.csv)

Voynich Manuscript changepoint plot code

multinomial\_changepoint\_voynich\_plot.r

library( tidyverse )

library( magrittr )

library( ggthemes )

library( showtext )

library( cowplot )

library( grimoire ) # https://github.com/weirddatascience/grimoire

# Fonts

font\_add( "voynich\_font", "resources/fonts/voynich/eva1.ttf")

font\_add( "main\_font", "resources/fonts/alchemy/1651 Alchemy/1651AlchemyNormal.otf")

font\_add( "bold\_font", "resources/fonts/alchemy/1651 Alchemy/1651AlchemyNormal.otf")

showtext\_auto()

mean\_changepoint\_prob <-

readRDS( "work/mean\_changepoint\_prob\_voynich.rds" )

changepoint\_plot <-

ggplot( mean\_changepoint\_prob ) +

geom\_col( aes( x=rowid, y=value ), fill=weird\_colours["blood"] ) +

labs( x="Folio", y="Probability of Changepoint" ) +

theme(

panel.background = element\_rect(fill = "transparent", colour = "transparent"),

plot.background = element\_rect(fill = "transparent", colour = "transparent"),

plot.title = element\_text( family="bold\_font", colour=weird\_colours["ink"], size=22 ),

plot.subtitle = element\_text( family="bold\_font", colour=weird\_colours["ink"], size=12 ),

axis.text = element\_text( family="bold\_font", colour=weird\_colours["ink"], size=12 ),

axis.title.x = element\_text( family="bold\_font", colour=weird\_colours["ink"], size=12 ),

axis.title.y = element\_text( family="bold\_font", colour=weird\_colours["ink"], angle=90, size=12 ),

axis.line = element\_line( colour=weird\_colours["ink"] ),

panel.grid.major.x = element\_blank(),

panel.grid.major.y = element\_blank(),

panel.grid.minor.x = element\_blank(),

panel.grid.minor.y = element\_blank()

)

# grimoire::decorate\_plot() from

parchment\_plot <-

decorate\_plot(

title="Voynich Folio Word-Length Changepoint",

subtitle="http://www.weirddatascience.net | @WeirdDataSci",

plot=changepoint\_plot,

bg\_image="resources/img/parchment.jpg",

footer="Data: http://www.voynich.nu" )

save\_plot("output/multinomial\_changepoint\_voynich\_plot.pdf",

parchment\_plot,

base\_width = 16,

base\_height = 9,

base\_aspect\_ratio = 1.78 )

Our model appears to identify a potential changepoint in the frequency of short words in the Voynich Manuscript somewhere around Folio 33.

In contrast to the analysis of *Tirant lo Blanc*, we have no *a priori* suspicion that multiple authors were involved in the creation of the Voynich Manuscript. As such, we may hypothesise this changepoint as reflecting a simple stylistic shift, a shift in content, or, as with the previous analysis, a shift in authorship.

**Mutatis Mutandis**

Recalling the topic model, and the manually-assigned topics hinted at by the diagrams in the manuscript, Folio 33 might not immediately arouse our suspicions as the most obvious candidate for such a shift in style, falling as it does some way through the manually-identified herbal section, and without an immediately apparent shift in the distribution of topics.

For simplicity of presentation and analysis we will work with the alternative 12-topic model suggested by the metrics in the previous post rather than the 34-topic model initially given there. The distribution of topics in this model can be presented as those of the 34-topic model were in our previous post.

Voynich Manuscript Folio Topic Heatmap |

We might, then, ask whether the distribution of topics from this model itself has a changepoint. Having interrogated the distribution of word length frequencies, we can pose the same questions to the distribution of assignments produced by the topic model. Similarly to above: were we to conceive of the topic model assignment for each page as being the result of the roll of some biased die, is there a notable point in the document where the bias of that die seems to shift?

Framed as such, there is little more required to apply this model to the topic assignments. Our Stan model, dredged from its slumber, merely needs to be provided with the topic model folio assignment data.

Voynich Manuscript topic model changepoint |

Voynich Manuscript topic model changepoint code

multinomial\_changepoint\_voynich\_topics.r

library( tidyverse )

library( tidyselect )

library( magrittr )

library( rstan )

library( tidytext )

library( gtools ) # (Specifically for `mixedsort` to sort column names numerically.

# Number of topics in model

num\_topics <- 12

# Load Voynich topic model data

message( "Reading Voynich topic model data..." )

voynich\_tbl <-

readRDS( paste0( "work/topic\_identity-", num\_topics, ".rds" )) %>%

select( -c( "gamma", "section" ) ) %>%

ungroup

# Pivot the data wider to be presented to Stan as a matrix of samples from a multinomial.

voynich\_lengths <-

voynich\_tbl %>%

mutate( count=1 ) %>%

pivot\_wider( names\_from = topic, values\_from = "count" ) %>%

select( -document ) %>%

select(mixedsort(peek\_vars())) %>%

replace( is.na(.), 0 )

topic\_fit\_file <- paste0( "work/multinomial\_changepoint\_voynich\_topic\_fit-", num\_topics, ".rds" )

if( not( file.exists( topic\_fit\_file ) ) ) {

message( "Fitting multinomial model.")

voynich\_topic\_multinom\_fit <-

stan( "multinomial\_changepoint.stan",

data=list(

num\_obs=226,

num\_cats=num\_topics,

y = as.matrix( voynich\_lengths ),

alpha = rep( 1, num\_topics ) ),

iter=8000,

seed=19300319,

control=list( adapt\_delta=0.9 ) )

saveRDS( voynich\_multinom\_fit, topic\_fit\_file )

} else {

message( "Loading saved multinomial model.")

voynich\_multinom\_fit <- readRDS( topic\_fit\_file )

}

# Extract the calculated changepoint probabilities to a simplex.

mean\_changepoint\_prob <-

extract( voynich\_multinom\_fit )$changepoint\_simplex %>%

as\_tibble( .name\_repair="unique" ) %>%

summarise\_all( mean ) %>%

pivot\_longer( everything() ) %>%

rowid\_to\_column()

# Save values for plotting

saveRDS( mean\_changepoint\_prob, file=paste0( "work/mean\_changepoint\_prob\_voynich\_topic-", num\_topics, ".rds" ) )

Voynich Manuscript topic model changepoint plot code

multinomial\_changepoint\_voynich\_topics\_plot.r

library( tidyverse )

library( magrittr )

library( ggthemes )

library( showtext )

library( cowplot )

library( grimoire )

# Fonts

font\_add( "voynich\_font", "resources/fonts/voynich/eva1.ttf")

font\_add( "main\_font", "resources/fonts/alchemy/1651 Alchemy/1651AlchemyNormal.otf")

font\_add( "bold\_font", "resources/fonts/alchemy/1651 Alchemy/1651AlchemyNormal.otf")

showtext\_auto()

# Specify the number of topics in the model

num\_topics <- 12

mean\_changepoint\_prob <-

readRDS( paste0("work/mean\_changepoint\_prob\_voynich\_topic-", num\_topics, ".rds" ) )

changepoint\_plot <-

ggplot( mean\_changepoint\_prob ) +

geom\_col( aes( x=rowid, y=value ), fill=weird\_colours["blood"] ) +

labs( x="Folio", y="Probability of Changepoint" ) +

theme(

panel.background = element\_rect(fill = "transparent", colour = "transparent"),

plot.background = element\_rect(fill = "transparent", colour = "transparent"),

plot.title = element\_text( family="bold\_font", colour=weird\_colours["ink"], size=22 ),

plot.subtitle = element\_text( family="bold\_font", colour=weird\_colours["ink"], size=12 ),

axis.text = element\_text( family="bold\_font", colour=weird\_colours["ink"], size=12 ),

axis.title.x = element\_text( family="bold\_font", colour=weird\_colours["ink"], size=12 ),

axis.title.y = element\_text( family="bold\_font", colour=weird\_colours["ink"], angle=90, size=12 ),

axis.line = element\_line( colour=weird\_colours["ink"] ),

panel.grid.major.x = element\_blank(),

panel.grid.major.y = element\_blank(),

panel.grid.minor.x = element\_blank(),

panel.grid.minor.y = element\_blank()

)

# grimoire::decorate\_plot() from

parchment\_plot <-

decorate\_plot(

title=paste0( "Voynich Folio Changepoint Probability - Topic Model (", num\_topics, " topics)"),

subtitle="http://www.weirddatascience.net | @WeirdDataSci",

plot=changepoint\_plot,

bg\_image="resources/img/parchment.jpg",

footer="Data: http://www.voynich.nu",

rel\_heights=c(0.1, 1, 0.05 ))

save\_plot( paste0("output/multinomial\_changepoint\_voynich\_topic-", num\_topics, "\_plot.pdf" ),

parchment\_plot,

base\_width = 16,

base\_height = 9,

base\_aspect\_ratio = 1.78 )

The location suggested for the topic model changepoint are surprisingly close to the results of the changepoint for the frequency counts of short words in the Voynich Manuscript. Whilst the changepoint probabilities are somewhat more diffuse in the topic model analysis, the most significant probability mass is centered around Folio 38, with much lower spikes extending out as far as Folio 55 and Folio 30.

In addition to the mutual support that these two analyses provide, it is notable that the major changpoint identified by both falls directly in the earlier portion of the first, major “herbal” section identified manually by Voynich scholars through inspection of the images accompanying the text. This suggests that that first section, at least in terms of textual content, is not as homogeneous as has previously been suggested. Future scholars investigating the structure of the Voynich Manuscript may therefore wish to direct more attention towards the earlier middle of the herbal section, around folios 30 to 40, to identify what dreadful changes may emerge at that point in the text.

We might naturally, but do not here, extend this analysis by shattering further the smooth unity of the Manuscript according to *multiple* changepoints. As such, we leave this analysis, and the means to conduct it more efficiently, for the dim future.

**Omnia mutantur, nihil interit**

Our analysis has provided us with an abstracted location in the text at which our unsettling suspicions of change lie. The two folios arousing greatest curiosity are therefore Folio 33 and Folio 38, which we present here to reify our horror. It is worth highlighting, however, that the changepoint analysis says nothing specific about these two folios; the model identifies that the inexplicable scrawling prior to the changepoint differs significantly from the maddeningly incomprehensible glyphs following it, nothing more.

The statistical properties that we have uncovered in the Voynich Manuscript over the past four posts reveal something of its inner structure. It supports, but cannot prove, that the Manuscript is not a hoax, and that it the text is most likely drawn from some natural language.

The changepoint analyses in this post are a powerful tool for identifying evolution and mutation in data, and the demonstrated example of stylometric analysis to Martorel’s *Tirant lo Blanc* support their use in revealing points of fracture in texts, without reference to the source language.

With specific relevant to the Voynich Manuscript, both the word frequency changepoint and the topic model changepoint suggest that the manscript’s contents shift significantly at some point around Folios 30 to 40. Given the previous assignment of topics based on manual identification of images accompanying the text, this presents a new avenue of investigation for Voynich researchers.

We have resisted that tantalising draw of attempts to translate the Voynich Manuscript. The tools we have applied are more broadly statistical and aim at unveiling structures and revealing patterns in the text; whilst they may provide information towards deciphering the text, that particular conundrum is for the future.

There are, as we would always wish, many avenues left unexplored in this particular labyrinth. The topic model is crude, and more subtle disassemblies could well provide a more refined view. The word frequency patterns support natural language, but we have not made any effort to correlate them with known languages. We have treated words as a unit of analysis, but have not looked in detail at the structure of likely prefixes and suffixes; similar words with differing endings are particularly notable in the topic model, and analysis of these could reveal much more than we have dared to attempt.

There could be much to learn from assessing multiple changepoints in the Manuscript. The presentation of the volume certainly supports its composition of multiple disparate sections; perhaps identifying an inexorable sequence of stylistic shifts could unveil still more of this structure.